PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G10L 5/04 | A1

(11) International Publication Number: WO 96/27870

(43) International Publication Date: 12 September 1996 (12.09.96)

(21) International Application Number: PCT/GB96/00529

(22) International Filing Date: 7 March 1996 (07.03.96)

(30) Priority Data:
95301478.4 7 March 1995 (07.03.95) EP
(34) Countries for which the regional or international application was filed: GB et al.

(71) Applicant *(for all designated States except US)*: BRITISH TELECOMMUNICATIONS PUBLIC LIMITED COMPANY [GB/GB]; 81 Newgate Street, London EC1A 7AJ (GB).

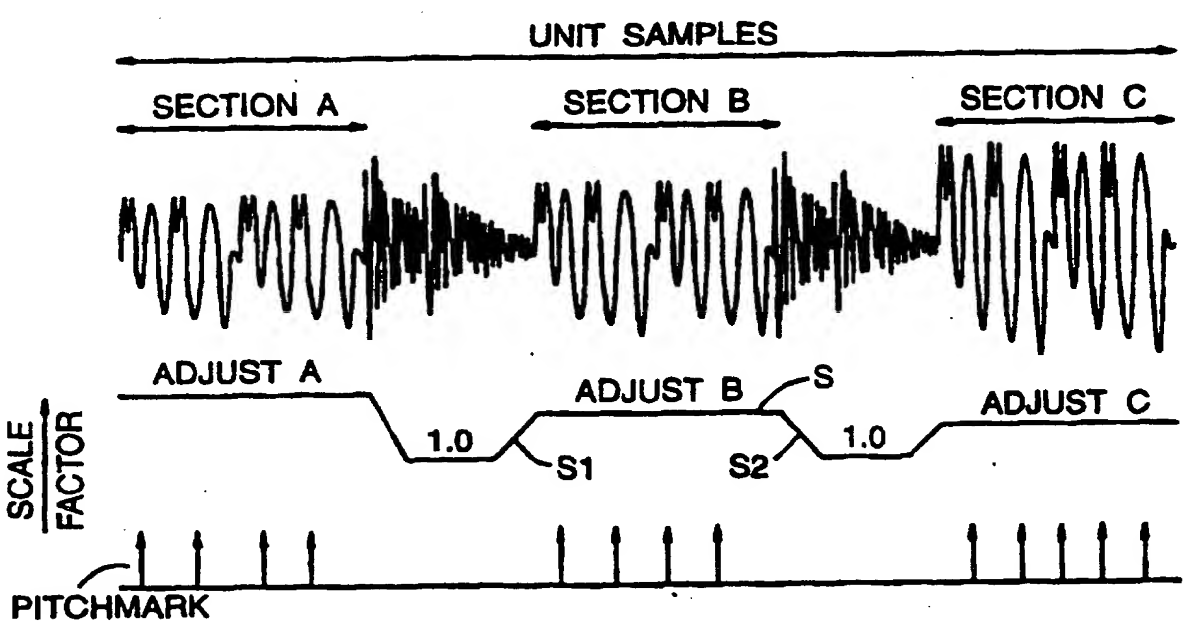
(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: LOWRY, Andrew [GB/GB]; 27 Ranelagh Road, Ipswich, Suffolk IP2 0AD (GB). BREEN, Andrew [GB/GB]; 50 Westerfield Road, Ipswich, Suffolk IP4 2UT (GB). JACKSON, Peter [GB/GB]; 36 Manor Road, Martlesham Heath, Ipswich, Suffolk IP5 7SY (GB).

(74) Agent: LLOYD, Barry, George, William; BT Group Legal Services, Intellectual Property Dept., 8th floor, 120 Holborn, London EC1N 2TE (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: SPEECH SYNTHESIS



(57) Abstract

Portions of recorded speech waveform (e.g. corresponding to phonemes) are combined to synthesise words. In order to provide a smoother delivery, each voiced portion of a waveform portion has its amplitude adjusted to a predetermined reference level. The scaling factor used is varied gradually over a transition region between such portions and between voiced and unvoiced portions.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
CS	Czechoslovakia	LT	Lithuania	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

SPEECH SYNTHESIS

One method of synthesising speech involves the concatenation of small units of speech in the time domain. Thus representations of speech waveform may be stored, and small units such as phonemes, diphones or triphones - i.e. units of less than a word - selected according to the speech that is to be synthesised, and concatenated. Following concatenation, known techniques may be employed to adjust the composite waveform to ensure continuity of pitch and signal phase. However, another factor affecting the perceived quality of the resulting synthesised speech is the amplitude of the units; preprocessing of the waveforms - i.e. adjustment of amplitude prior to storage - is not found to solve this problem, inter alia because the length of the units extracted from the stored data may vary.

According to the present invention there is provided a speech synthesiser comprising

- a store containing representations of speech waveform;

- selection means responsive in operation to phonetic representations input thereto of desired sounds to select from the store units of speech waveform representing portions of words corresponding to the desired sounds;

- means for concatenating the selected units of speech waveform characterised by means for adjusting the amplitude of at least the voiced portion relative to a predetermined reference level.

One example of the invention will now be described, by way of example, with reference to the accompanying drawings, in which:

Figure 1 is a block diagram of one example of speech synthesis according to the invention;

Figure 2 is a flow chart illustrating operation of the synthesis; and

Figure 3 is a timing diagram.

In the speech synthesiser of Figure 1, a store 1 contains speech waveform sections generated from a digitised passage of speech, originally recorded by a human speaker reading a passage (of perhaps 200 sentences) selected to contain all possible (or at least, a wide selection of) different sounds.

Accompanying each section is stored data defining "pitchmarks" indicative of points of glottal closure in the signal, generated in conventional manner during the original recording.

An input signal representing speech to be synthesised, in the form of a phonetic representation is supplied to an input 2. This input may if wished be generated from a text input by conventional means (not shown). This input is processed in known manner by a selection unit 3 which determines, for each unit of the input, the addresses in the store 1 of a stored waveform section corresponding to the sound represented by the unit. The unit may, as mentioned above, be a phoneme, diphone, triphone or other sub-word unit, and in general the length of a unit may vary according to the availability in the waveform store of a corresponding waveform section.

The units, once read out, are concatenated at 4 and the concatenated waveform subjected to any desired pitch adjustments at 5.

Prior to this concatenation, each unit is individually subjected to an amplitude normalisation process in an amplitude adjustment unit 6 whose operation will now be described in more detail. The basic objective is to normalise each voiced portion of the unit to a fixed RMS level before any further processing is applied. A label representing the unit selected allows the reference level store 8 to determine the appropriate RMS level to be used in the normalisation process. Unvoiced portions are not adjusted, but the transitions between voiced and unvoiced portions may be smoothed to avoid sharp discontinuities. The motivation for this approach lies in the operation of the unit selection and concatenation procedures. The units selected are variable in length, and in the context from which they are taken. This makes preprocessing difficult, as the length, context and voicing characteristics of adjoining units affect the merging algorithm, and hence the variation of amplitude across the join. This information is only known at run-time as each unit is selected. Postprocessing after the merge is equally difficult.

The first task of the amplitude adjustment unit is to identify the voiced portions(s) (if any) of the unit. This is done with the aid of a voicing detector 7 which makes use of the pitch timing marks indicative of points of glottal closure in

the signal, the distance between successive marks determining the fundamental frequency of the signal. The data (from the waveform store 1) representing the timing of the pitch marks are received by the voicing detector 7 which, by reference to a maximum separation corresponding to the lowest expected fundamental frequency, identifies voiced portions of the unit by deeming a succession of pitch marks separated by less than this maximum to constitute a voiced portion. A voiced portion whose first (or last) pitchmark is within this maximum of the beginning (or end) of the speech unit is, respectively, considered to begin at the beginning of the unit or end at the end of the unit. This identification step is shown as step 10 in the flowchart shown in Figure 2.

The amplitude adjustment unit 6 then computes (step 11) the RMS value of the waveform over the voiced portion, for example the portion B shown in the timing diagram of Figure 3, and a scale factor S equal to a fixed reference value divided by this RMS value. The fixed reference value may be the same for all speech portions, or more than one reference value may be used specific to particular subsets of speech portions. For example, different phonemes may be allocated different reference values. If the voiced portion occurs across the boundary between two different subsets, then the scale factor S can be calculated as a weighted sum of each fixed reference value divided by the RMS value. Appropriate weights are calculated according to the proportion of the voiced portion which falls within each subset. All sample values within the voiced portion are (step 12 of Figure 2) multiplied by the scale factor S. In order to smooth voiced/unvoiced transitions, the last 10ms of unvoiced speech samples prior to the voiced portion are multiplied (step 13) by a factor S_1 which varies linearly from 1 to S over this period. Similarly, the first 10ms of unvoiced speech samples following the voiced portion are multiplied (step 14) by a factor S_2 which varies linearly from S to 1. Tests 15, 16 in the flowchart ensure that these steps are not performed when the voiced portion respectively starts or ends at the unit boundary.

Figure 3 shows the scaling procedure for a unit with three voiced portions A, B, C, D, separated by unvoiced portions. Portion A is at the start of the unit,

so it has no ramp-in segment, but has a ramp-out segment. Portion B begins and ends within the unit, so it has a ramp-in and ramp-out segment. Portion C starts within the unit, but continues to the end of the unit, so it has a ramp-in, but no ramp-out segment.

This scaling process is understood to be applied to each voiced portion in turn, if more than one is found.

Although the amplitude adjustment unit may be realised in dedicated hardware, preferably it is formed by a stored program controlled processor operating in accordance with the flowchart of Figure 2.

CLAIMS

1. A speech synthesiser comprising

- a store containing representations of speech waveform;

- selection means responsive in operation to phonetic representations input thereto of desired sounds to select from the store units of speech waveform representing portions of words corresponding to the desired sounds;

- means for identifying voiced portions of the selected units

- means for concatenating the selected units of speech waveform; characterised by means arranged to adjust the amplitude of the voiced portions of the units relative to a predetermined reference level and to leave unchanged at least part of any unvoiced portion of the unit.

2. A speech synthesiser according to Claim 1 in which the adjusting means is arranged to scale the or each voiced portion by a respective scaling factor, and to scale the adjacent part of any abutting unvoiced portion by a factor which varies monotonically over the duration of that part between the scaling factor and unity.

3. A speech synthesiser according to Claim 1 or 2 in which a plurality of reference levels is used, the adjusting means being arranged for each voiced portion, to select a reference level in dependence upon the sound represented by that portion.

4. A speech synthesiser according to Claim 3 in which each phoneme is assigned a reference level and any voiced portion containing waveform segments from more than one phoneme is assigned a reference level which is a weighted sum of the levels assigned to the phonemes contained therein, weighted according to the relative durations of the segments.

Fig.1.

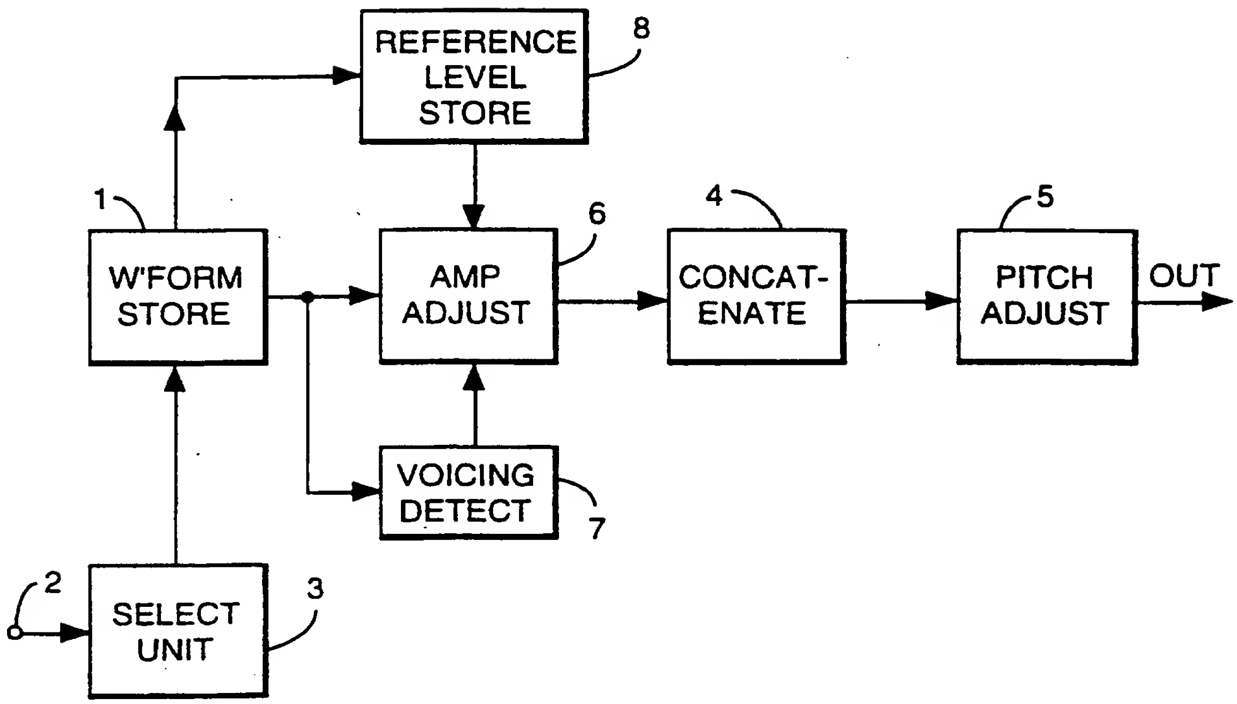

REFERENCE LEVEL STORE
8
1
W'FORM STORE
6
AMP ADJUST
4
CONCAT-ENATE
5
PITCH ADJUST
OUT
VOICING DETECT
7
2
SELECT UNIT
3

Fig.3.

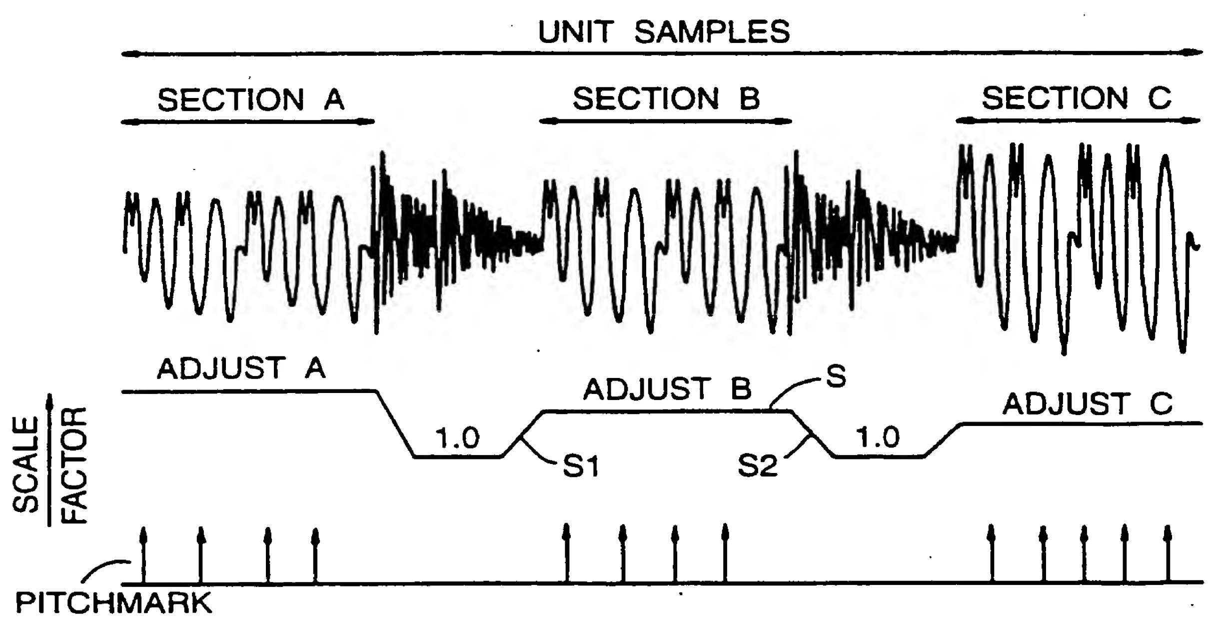
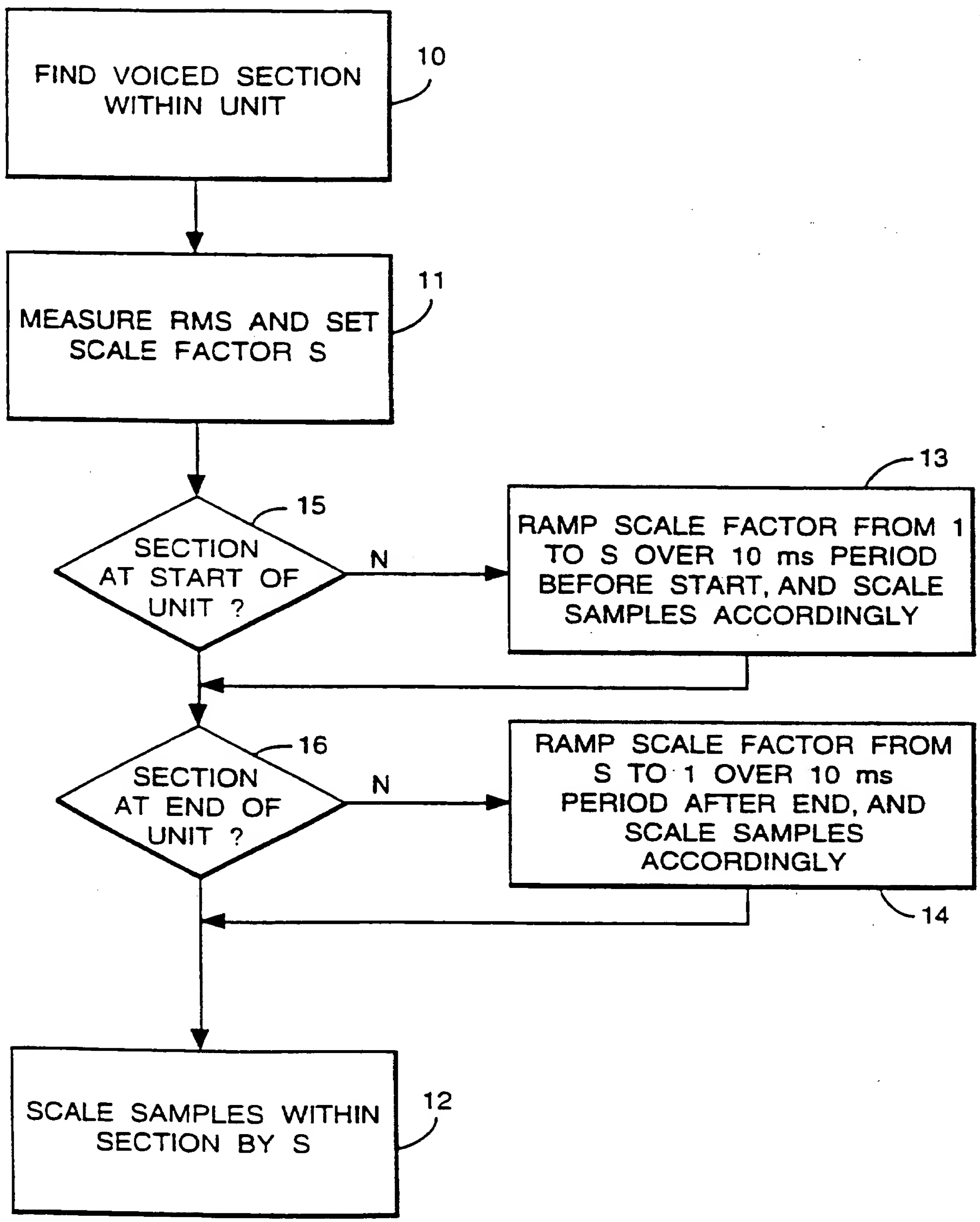

UNIT SAMPLES
SECTION A
SECTION B
SECTION C
ADJUST A
ADJUST B
ADJUST C
S
1.0
1.0
S1
S2
SCALE FACTOR
PITCHMARK

Fig.2.


FIND VOICED SECTION WITHIN UNIT
10
MEASURE RMS AND SET SCALE FACTOR S
11
SECTION AT START OF UNIT ?
15
N
RAMP SCALE FACTOR FROM 1 TO S OVER 10 ms PERIOD BEFORE START, AND SCALE SAMPLES ACCORDINGLY
13
SECTION AT END OF UNIT ?
16
N
RAMP SCALE FACTOR FROM S TO 1 OVER 10 ms PERIOD AFTER END, AND SCALE SAMPLES ACCORDINGLY
14
SCALE SAMPLES WITHIN SECTION BY S
12

INTERNATIONAL SEARCH REPORT

Inte nal Application No
PCT/GB 96/00529

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 G10L5/04

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP,A,0 107 945 (TOKYO SHIBAURA ELECTRIC) 9 May 1984 see page 2, line 6 - page 3, line 23; figures 2-4 ---	1
A	DE,A,19 22 170 (NIPPON TELEGRAPH & TELEPHONE) 13 November 1969 see page 8; figure 2B ---	1
A	JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, vol. 66, no. 5, November 1979, NEW YORK, US, pages 1325-1332, XP000567943 SHADLE ET AL.: "Speech synthesis by linear interpolation of spectral parameters between dyad boundaries" see paragraph I; figure 1 ---	1

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

2

Date of the actual completion of the international search	Date of mailing of the international search report
17 June 1996	21.06.96

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Lange, J

Inte   nal Application No
PC1/GB 96/00529

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING 86, 7 - 11 April 1986, TOKYO, JP, pages 2019-2022 vol.3, XP000567953 YAZU ET AL.: "The speech synthesis system for an unlimited Japanese vocabulary" see paragraph 3.3; figure 4 ---	1
A	EP,A,0 427 485 (CANON) 15 May 1991 see page 6, line 35 - page 7, line 27; figures 5,6 -----	1

2

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/GB 96/00529

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP-A-0107945	09-05-84	JP-A- 59072494	24-04-84
DE-A-1922170	13-11-69	FR-A- 2007620	09-01-70
		GB-A- 1224137	03-03-71
		SE-B- 355887	07-05-73
EP-A-0427485	15-05-91	JP-A- 3149600	26-06-91
		JP-A- 3203798	05-09-91
		JP-A- 3203799	05-09-91
		JP-A- 3203793	05-09-91
		JP-A- 3203800	05-09-91
		JP-A- 3198098	29-08-91
		US-A- 5220629	15-06-93